

RESEARCH

Open Access



Information access equality on generative models of complex networks

Xindi Wang^{1*}, Onur Varol² and Tina Eliassi-Rad^{1,3}

*Correspondence:
xindi.w1993@gmail.com

¹ Network Science Institute,
Northeastern University, Boston,
USA

² Faculty of Engineering
and Natural Sciences, Sabanci
University, Istanbul, Turkey

³ Khoury College of Computer
Sciences, Northeastern
University, Boston, USA

Abstract

It is well known that networks generated by common mechanisms such as preferential attachment and homophily can disadvantage the minority group by limiting their ability to establish links with the majority group. This has the effect of limiting minority nodes' access to information. We present the results of an empirical study on the equality of information access in network models with different growth mechanisms and spreading processes. For growth mechanisms, we focus on the majority/minority dichotomy, homophily, preferential attachment, and diversity. For spreading processes, we investigate simple versus complex contagions, different transmission rates within and between groups, and various seeding conditions. We observe two phenomena. First, information access equality is a complex interplay between network structures and the spreading processes. Second, there is a trade-off between equality and efficiency of information access under certain circumstances (e.g., when inter-group edges are low and information transmits asymmetrically). Our findings can be used to make recommendations for mechanistic design of social networks with information access equality.

Keywords: Inequality, Information access, Information diffusion, Complex networks

Introduction

The initial hopes that an increasingly interconnected world will provide more equal opportunities and remove information barriers are dampened by observations of echo chambers and polarisation on social networks (Adamic and Glance 2005; Iyengar and Hahn 2009; Prior 2007; Garrett 2009; Gentzkow and Shapiro 2011; Barberá et al. 2015). Recent studies have also found that discrimination in social networks can arise through simple mechanisms. For example, homophily and minority group size can lead to glass-ceiling (Avin et al. 2015; Stoica et al. 2018) and chasm effects (Zhang et al. 2021), influence the ranking of minority nodes (Karimi et al. 2018; Oliveira et al. 2021), and create perception bias in social networks (Lee et al. 2019). In addition, a user's popularity and prestige on social networking platforms can be altered through the use of automation and social bots to gain unfair advantages (Aiello et al. 2012; Messias et al. 2013; Woolley 2016; Varol et al. 2017; Varol and Uluturk 2020).

An important role of networks is that they disseminate information, from job and business opportunities (Boyd et al. 2014) to medical resources that can be accessed (Freese

and Lutfey 2011), and such information is critical to people's lives. However, inequality in access to information in a network has not been fully explored. This is a topic that has been discussed in the social and political sciences. For instance, economically disadvantaged people have less access to new technologies such as the Internet (DiMaggio and Cohen 2021), and groups starting with worse employment status are more likely to experience persistent unemployment in the labor market (Calvo-Armengol and Jackson 2004).

A research topic closely related to information dissemination in networks is information maximization, which aims to find the optimal starting (or seed) nodes in a network that maximize information spread (Aral and Walker 2011; Kempe et al. 2003). This problem has been studied over the years, but the fairness aspect in this problem is relatively new. For example, Stoica and Chaintreau (Stoica and Chaintreau 2019) investigate two fairness criteria for information maximization: fairness for early adopters (where seeds should be proportional to the group population) and fairness in outreach (where the information should reach the same percentage of each group in the population). They experiment on a social network extracted from Instagram. In subsequent work, Stoica et al. (2020) proposed diversity-enhancing interventions in seeding and explored the complicated relationship between diversity and efficiency. Others have explored how fairness constraints can be incorporated into the information maximization algorithm, including the maximin constraint¹ (Fish et al. 2019; Becker et al. 2021; Tsang et al. 2019), diversity constraint (Tsang et al. 2019), welfare theory (Rahmattalabi et al. 2021), and time-sensitive constraint on information (Ali et al. 2019). Prior literature sheds some light on information equality in networks, but it mostly considers the problem as an algorithmic problem (given a fixed graph, what are the optimal information seeds), rather than a characteristic problem (what kind of network can better promote information equality). Moreover, the spreading process considered is mostly the independent cascade model, which may not be representative of real-world processes. Different from previous works, Jalali et al. (2020) define a criterion for information unfairness (based on whether information flows equally among all groups in a network) and present an algorithm that adds edges to the network to reduce the information unfairness criterion. They experiment on a social network extracted from DBLP. This study does not provide much information about the performance of various spreading processes on different networks. In a recent tutorial, Venkatasubramanian et al. (2021) survey recent work on fairness in networks. They cover topics such as social capital, information access (the subject of our work), and interventions. In lieu of a survey paper on fairness in networks, we recommend their tutorial to newcomers in this topic.

Here, we study the problem of information access equality in networks from a characteristic point of view. Using several network models with two mutually exclusive groups in the population (majority vs. minority), we generate networks with different properties and constraints that are representative of mechanisms in real networks. We define information access equality as follows: *for a given process and seeds, the majority and minority nodes should receive information at similar rates across various stages of the*

¹ The maximin constraint requires that the least advantageous group should be the most improved during the optimization process.

spreading processes. Our goal is to provide insight into which network characteristics may affect information access equality and in what ways, and to make recommendations for systematic mechanism design. We find that, in general, homophily and preferential attachment can harm information access equality, while introducing diversity can promote information access equality. However, too much diversity can affect the efficiency of information spreading. More importantly, we find that information access equality depends not only on the network, but also on the characteristics of the spreading process.

Our main contributions are as follows: (1) Information access equality is a complex interplay between network structure and the spreading process. (2) There is a trade-off between equality and efficiency of information access under certain circumstances (e.g., when the network has low inter-group edges and asymmetric transmission). (3) Spreading process features are statistically significant (p -value ≤ 0.05) when it comes to information access equality. A symmetric transmission rate will increase equality compared to an asymmetric one. A simple contagion will decrease equality compared to a complex contagion. Compared to a mid seeding portion, a low minority seeding portion will decrease equality and a high minority seeding portion will increase equality. (4) Network features are *not* always statistically significant. However, two network features stand out with respect to information access equality: degree inequality (as measured by power inequality or moment glass ceiling) and network distance (as measured by shortest path or diameter). Our findings can be used to recommend connections that steer an online social network toward information access equality. Ideally, such recommendation systems will first classify the spreading process (simple vs. complex contagion, symmetric vs. asymmetric transmission), then based on that classification recommend new connections to their users.

Network models

Our study includes several generative models for complex networks. Each model generates a network with two mutually exclusive groups of nodes: majority and minority. Let m denote the proportion of minority nodes. The majority/minority dichotomy is based on population size: minority nodes make up less than 50% of the network (i.e., $m < 0.5$). A real-world example of such a dichotomy is a computer science collaboration network where male scientists are in the majority and female scientists are in the minority.

The generative models produce undirected, unweighted networks. Starting from an initial network, a new node enters the network at each time step and connects to some nodes according to a mechanism mandated by the model. The initialization procedure is the same across the models: a node from the majority group connects to a node from the minority group. Table 1 summarizes each model's properties and provides a comparison between the proposed models.

Random network

In the Random Network model, at each time step a new node enters with probability m as a minority and $1 - m$ as a majority. With uniform probability, each node connects to l existing nodes. This is a growing version of the Erdős-Rényi Random Network (Erdős and Rényi 1960).

Table 1 Relationship between the network models.

Random network	Preferential attachment	Homophily	Diversity	Resultant network
✓	✓			BA
✓		✓		Random Homophily
✓	✓	✓		Homophily BA
✓		✓	✓	Diversified Homophily
✓	✓	✓	✓	Diversified Homophily BA

All network models are based on Random Network, a growing network adaptation of the Erdős-Rényi random network (Erdős and Rényi 1960). We focus on three mechanisms: preferential attachment, homophily, and diversity (through adding inter-group edges). Starting from a Random Network and adding preferential attachment results in the BA model. Starting from a Random Network and adding homophily results in the Random Homophily model. Starting from a Random Network and adding both preferential attachment and homophily results in the Homophily BA model. Starting from Random Homophily and adding diversity results in the Diversified Homophily model. Starting from Homophily BA and adding diversity gives us Diversified Homophily BA model. In our study, we treat BA and Random Homophily as the variations of the Homophily BA model, and Diversified Homophily as a variation of the Diversified Homophily BA

Homophily BA

In the Homophily BA model (Avin et al. 2015; Karimi et al. 2018; Lee et al. 2019),² two ingredients influence link formation: preferential attachment and homophily. The parameters of Homophily BA include: (1) the proportion of minority nodes m , (2) the number of edges l for each new node, (3) the homophily matrix H with entries $h_{g_i g_j}$, and (4) the preferential attachment strength α . Group memberships are denoted by g_i and g_j for nodes i and j , respectively. Each node i can either belong to the majority (maj) or the minority (min) group. Thus, we have the following homophily matrix H :

$$H = \begin{bmatrix} h_{maj,maj} & h_{maj,min} \\ h_{min,maj} & h_{min,min} \end{bmatrix}$$

For simplicity, we assume H is a doubly stochastic matrix, with $h_{maj,min} = h_{min,maj}$ and $h_{maj,maj} = h_{min,min}$. We define $h = h_{maj,maj} = h_{min,min}$. Our assumption reduces the homophily matrix to $H = \begin{bmatrix} h & 1-h \\ 1-h & h \end{bmatrix}$. When $h = 1$, the network is perfectly homophilic – i.e., the network has two distinct groups: majority and minority. The groups are connected with a single inter-group edge (due to network initialization). On the other hand, if $h = 0$, the network is perfectly heterophilic. It has been found that homophily can be different for different groups (Messias et al. 2017; Karimi et al. 2018). We leave the exploration of asymmetric homophily for future work.

The Homophily BA networks grow as follows.

- At each time step, a new node j enters with probability m as a minority and $1 - m$ as a majority. Let g_j denote its group membership.
- Node j connects to l nodes. Each connection to node i in the network is made with probability Π_i :

² BA is short for Barabási-Albert (Albert and Barabási 2002).

$$\Pi_i = \frac{h_{g_j g_i} d_i^\alpha}{\sum_i h_{g_j g_i} d_i^\alpha}, \tag{1}$$

where d_i is the degree of node i .

We are also interested in two special cases of Homophily BA: (1) Without preferential attachment (Random Homophily with $\alpha = 0$, which is equivalent to a Stochastic Block Model (Holland et al. 1983)) and (2) With random mixing (BA (Albert and Barabási 2002) with $h = 0.5$).

Diversified homophily BA

We propose the Diversified Homophily BA model to encourage inter-group connections while maintaining some degree of homophily. The parameters of the Diversified Homophily BA include: (1) The proportion of minority nodes m , (2) The number of edges l for each new node, (3) The homophily matrix H (with the same assumption as in Homophily BA), (4) Preferential attachment strength α , (5) The number of diversified edges for each node l_d , and (6) The diversification probability p_d . The Diversified Homophily BA networks grow as follows:

- At each time step, a new node j enters with probability m as a minority and $1 - m$ as a majority. Let g_j denote its group membership.
- Node j forms $l - l_d$ links following Homophily BA mechanism $\Pi_i = \frac{h_{g_j g_i} k_i^\alpha}{\sum_i h_{g_j g_i} k_i^\alpha}$. Let S_j denote the nodes connected at this step.
- Create diversified links for node j . Let p_d denote the probability for node j to connect to a node in the opposite group. Thus, the probability that node j connects to node k is defined as follows:

$$p_{jk} = \begin{cases} p_d, & g_j \neq g_k \\ 1 - p_d, & g_j = g_k. \end{cases} \tag{2}$$

Let N_{S_j} denote the neighboring nodes for all nodes in S_j . Recall that S_j is the set of nodes to which the new node j connected via the Homophily BA mechanism. Generate a total of l_d diversified links by connecting node j to the nodes in N_{S_j} . This process is as follows. Node j connects to each node $k \in N_{S_j}$ with probability $\Pi_{jk} \propto p_{jk} \times \frac{1}{|d_k - d_j|}$, where d_i is the degree of node $i \in S_j$. The idea behind this step is to connect to nodes that are of opposite group, but with similar degree to existing neighbors.

Diversified Homophily BA is reducible to Homophily BA when the number of diversified links $l_d = 0$. We consider this model with $l_d \neq 0$ as a potential remedy for reducing Homophily BA's inequality, specifically, the inequality attributable to homophily (as opposed to preferential attachment). The inequality due to homophily has been documented before (Avin et al. 2015; Karimi et al. 2018; Lee et al. 2019). We are interested in a special case under Diversified Homophily BA with $\alpha = 0$, which removes preferential attachment. We call this case Diversified Homophily.

Information access equality

Currently, there is no consensus on how to measure the equality of information access in a network. In this work, we choose to measure it experimentally. That is, we simulate spreading processes on a given network and observe the differences in information spread among majority and minority nodes. In this way, we can explore how the type of contagion, the transmission rate, and the information seeds affect equality.

Varying processes that spread information

We consider several variations of dynamic processes. The first variation we consider is the distinction between *simple* vs. *complex contagion*. Previous studies have found that while simple contagion is appropriate for most diseases and spread of simple facts, complex contagion is more common for collective behaviors, such as the spread of new technologies and innovations, growth of social movements, and spread of misinformation (Centola and Macy 2007; Vespignani 2012; Mønsted et al. 2017; Granovetter 1978; Centola 2010; Anderson and May 1992; Daley and Kendall 1964; Romero et al. 2011; Weng et al. 2013; Shao et al. 2018). Same network properties may respond differently to these two types of contagion. For example, in simple contagion, weak ties are considered important because they can disseminate information to an isolated part of the network. However, in complex contagion, weak ties might not be as useful because one link is not sufficient to disseminate information (Granovetter 1973; Centola and Macy 2007). For modeling, we rely on a Susceptible-Infectious (SI) model (Barrat et al. 2008) for both simple and complex contagion. For complex contagion, we add the parameter of activation threshold a , where each node can only be infected if a portion of its neighborhood is infected.

The second variation we consider is the difference in *transmission rate* within and between groups. Although the transmission rate can be different at the node level (Aral and Dhillon 2018), we study the simple case where the transmission rate differs by group. In particular, we are interested in the case where the transmission rate between all nodes is the same (symmetric) and the case where the transmission rate between different groups is lower than within the same group (asymmetric). There are four sets of transmission rates: $r_{maj \rightarrow maj}$, $r_{min \rightarrow min}$, $r_{maj \rightarrow min}$ and $r_{min \rightarrow maj}$. Here we assume that $r_{maj \rightarrow maj} = r_{min \rightarrow min} = r_{within}$ and $r_{maj \rightarrow min} = r_{min \rightarrow maj} = r_{between}$. We simulate the SI process with symmetric transmission rate as $r_{within} = r_{between}$, and asymmetric transmission rate as $r_{within} > r_{between}$.

The last variation we consider is the *location where the information is seeded*. Different seeding locations can advantage or disadvantage a group. For example, imagine a network with two groups (majority vs. minority) connected by a single edge. Under this circumstance, if all seeds belong to the majority community, we would expect the minority group to be disadvantaged. To test the influence of seed location, we assume random seeding (as opposed to targeted seeding, such as selecting higher degree nodes or another form of seeding that assumes knowledge of node characteristics). We assume different portions of minority seeds: low (minority seeds below 30% of the seeded population), mid (minority seeds between 30% and 70% of the seeded population), and high (minority seeds above 70% of the seeded population). For brevity, we show only the

low and high minority seeding portions. The results for mid and high minority seeding portions are similar.³ Note that, in general, seeding a node with a piece of information comes at a cost. Thus, one can expect low seeding portions in real-life scenarios.

Measuring information access equality

To measure information access equality, we compare the number of nodes in state I (Infected) between the majority and minority groups. Note that the majority and minority groups are of different sizes. To allow a fair comparison, we define I_{maj} and I_{min} as the *fraction* of nodes in state I (Infected) of each group. In addition, since spreading processes can take different times in different networks, we normalize the I_{maj} and I_{min} by the length of the spreading process T to obtain $I_{maj}(t/T)$ and $I_{min}(t/T)$. Then, we calculate the relative difference between them as

$$\Delta I(t/T) = \frac{I_{maj}(t/T) - I_{min}(t/T)}{I_{maj}(t/T) + I_{min}(t/T)}.$$

We choose the denominator to be $I_{min}(t/T) + I_{maj}(t/T)$, which corresponds to twice the mean of $I_{maj}(t/T)$ and $I_{min}(t/T)$, making this metric symmetric and bounded between -1 and 1 . When there is equality in information access, $\Delta I(t/T) = 0$; and the closer $\Delta I(t/T)$ is to 0 , the greater the equality between the two groups. $\Delta I(t/T) > 0$ means that the majority group has a greater advantage. $\Delta I(t/T) < 0$ means that the minority group has a greater advantage. This measure is intuitive and allows us to inspect equality in information access at different stages of the spreading process.

Network measures

To understand the structure and properties of the network, and to gain insight into the possible roots of information access inequality, we examine several network measures. The first set captures homophily in terms of dyadicity and heterophilicity (Park and Barabási 2007). The second set computes distance in the network, which is directly related to information spread. The third set calculates difference in social capital (Burt 2000; Berlingerio et al. 2013) in terms of degree inequalities between the two groups.

Homophily: dyadicity and heterophilicity

The homophily parameter h in Homophily BA and Diversified Homophily BA may not reflect the actual homophily level of the final network. To measure the homophily effect more accurately, we calculate the dyadicity and heterophilicity score of the network (Park and Barabási 2007). Suppose we are given a network with N nodes and M edges. We have n_{maj} majority nodes and n_{min} minority nodes. Thus, $N = n_{maj} + n_{min}$. Let $p = \frac{2M}{N(N-1)}$ denote the average probability that two nodes are connected. If the group labels (majority or minority) are distributed randomly among the nodes, then the expected number of edges within the majority group is $\bar{m}_{maj,maj} = \binom{n_{maj}}{2}p$; the expected number of edges within the minority group is $\bar{m}_{min,min} = \binom{n_{min}}{2}p$; and the expected number of edges between the two groups is $\bar{m}_{maj,min} = \binom{n_{maj}}{1} \binom{n_{min}}{1} p$. Given the actual

³ The results for the mid minority seeding are in the Additional file 1.

number of edges $m_{maj,maj}$, $m_{min,min}$ and $m_{maj,min}$, dyadicity among the majority is $D_{maj} = \frac{m_{maj,maj}}{\bar{m}_{maj,maj}}$ and among the minority is $D_{min} = \frac{m_{min,min}}{\bar{m}_{min,min}}$; heterophilicity is $H = \frac{m_{maj,min}}{\bar{m}_{maj,min}}$. For a network with random mixing of among groups, dyadicity and heterophilicity are expected to be close to 1.

Distance: average shortest path length and diameter

We measure the average shortest path length and the diameter of the network. These two measures are directly related to the efficiency of information spreading. The higher the shortest path length or diameter, the longer it takes for information to spread to the entire network.

Social capital: degree equality

We are interested in whether there is a relationship between equality in social capital and information access. We use the following measures for degree equality. (1) *Earth Mover Distance (EMD)*. We calculate the Earth Mover Distance (EMD) between the degree distributions of the majority and minority nodes. The smaller the distance, the more equal the two groups are. (2) *Power Inequality (PI)*. Power inequality (Avin et al. 2015) measures the “power” (average degree) ratio of the minority and majority groups. Let \bar{d}_{min} and \bar{d}_{maj} denote the average degrees of the minority and majority nodes, respectively. The power inequality is defined as $PI = \frac{\bar{d}_{min}}{\bar{d}_{maj}}$. A network with power equality should have $PI = 1$; and the lower the PI , the more disadvantaged are the minority nodes. (3) *Moment Glass Ceiling (g)*. Moment glass ceiling measures the glass-ceiling effect in a network (Hymowitz and Schellhardt 1986; Avin et al. 2015). The intuition behind it is that a larger second moment (and assuming a similar average degree, i.e., no power inequality) leads to a larger variance in the distribution and thus a significantly larger number of nodes with high degree (i.e., hubs). Let $E(d_{min}^2)$ and $E(d_{maj}^2)$ denote the second moment of the degree distribution for the minority and majority nodes, respectively. The moment glass ceiling is defined as $g = \frac{E(d_{min}^2)}{E(d_{maj}^2)}$. A network with no glass-ceiling effect should have $g = 1$; and the lower the g , the more disadvantaged are the minority nodes.

Experiments

We report results on five sets of experiments. Experiment 1 is on information access equality across different network models. For each network model and each spreading process, we conduct 100K runs. Specifically, we repeat the following procedure 10 times: (1) We generate 100 networks for each model. (2) For each spreading process, we run 100 trials. Thus, we end up with a total of 100K ($= 10 \times 100 \times 100$) runs for each model and each spreading process. We observed no discernible effect on the results with 200 runs, 1K runs, and 10K runs. Experiment 2 is on the effects of different network parameters on information access equality. Experiment 3 is on information access equality in real-world networks. Experiment 4 is on regression analysis to detect significant features in information access equality. Experiment 5 is on the relationship between information equality and spreading efficiency.

For simplicity, we control the parameter space for the information spreading simulations. For all synthetic models, we set the number of seed nodes to be $s = 10$, which is 0.2% of the network, reflecting the assumption that the budget for seeding is limited. For the experiments involving real-world networks, we vary the seeding portions from 0.2% to 1% to 10%. In symmetric transmission, $r_{within} = r_{between} = 0.7$. In asymmetric transmission, $r_{within} = 0.7$ and $r_{between} = 0.3$. The activation threshold for complex contagion is $a = 0.1$. We did not perform experiments in which the size of the network (in terms of the number of nodes and edges) was varied. Our results are independent of the size of the network, except for the fact that the transmission process takes longer in larger networks.

Experiment 1: information access equality across different network models

We create networks with $N = 5000$ nodes, $m = 0.2$ (i.e., 20% of the nodes are minority nodes) and $l = 2$. We set $h = 0.8$ and $\alpha = 1$ in Homophily BA and Diversified Homophily BA, including their variations (namely, Random Homophily and Diversified BA). For Diversified Homophily BA, we set $l_d = 1$ and $p_d = 0.6$.

Network measures

In Fig. 1a, b, we compare the networks generated by the various models based on their degree distribution, statistics on nodes and edges, dyadicity, heterophilicity, average shortest path and diameter. For the degree distribution, we show the result of one sample network under each model. For all other statistics, we show the average of 100 realizations.

As expected, we observe that Random Network has dyadicity and heterophilicity of 1. For all models except Random Network, dyadicity in the minority group is larger than dyadicity in the majority group. Homophily BA, Random Homophily and Diversified Homophily have heterophilicity values below 1. This suggests that there are fewer edges between the groups than expected in random mixing. BA and Diversified Homophily BA have heterophilicity slightly above 1. This suggests that there are slightly more edges between the groups than expected in random mixing. As expected, the diversified versions of network models have larger heterophilicity compared to their undiversified versions.

We observe that Homophily BA and BA have the smallest average shortest path lengths and diameters due to the presence of hubs. Diversified Homophily and Diversified Homophily BA have the largest average shortest path length and diameter.

Figure 1c shows the social capital (degree) equality measures across different network models. The Earth Mover Distance between the minority degree distribution and the majority degree distribution is largest in Homophily BA and smallest in Random Network. For Power Inequality and Moment Glass Ceiling, the minority nodes have the most advantage in Diversified Homophily BA, while the majority nodes have the most advantage in Homophily BA.

Information access equality

In Fig. 2, we plot information access equality as a heatmap, where the x-axis represents the different stages of the spreading process t/T , and the y-axis is the different models.



Fig. 1 Statistics of network models. All networks have $N = 5000$ nodes, where each node joins the network with $l = 2$ edges and 20% of the nodes are minority. We set $h = 0.8$, $e_o = 1$, $\alpha = 1$ and $p_d = 0.6$. **a** Degree distributions. BA, Homophily BA, and Diversified Homophily BA all heavy-tailed degree distributions because of preferential attachment. **b** Basic Statistics. As expected, Random Network has dyadicity and heterophilicity of 1. For all network models expect Random Network, dyadicity in the minority group is larger than dyadicity in the majority group. The diversified versions of network models have larger heterophilicity compared to their undiversified versions because of the inter-group edges in the diversified versions. Homophily BA and BA have the smallest average shortest path lengths and diameters due to the presence of hubs. Diversified Homophily and Diversified Homophily BA have the largest average shortest path length and diameter. **c** Social Capital Equality Measures. For each equality measure, we sort the generative network models from most equal (left) to least equal (right). The earth mover distance is the distance between the degree distributions of the minority and majority groups. Homophily BA has the largest earth mover distance, while Random Network has the smallest distance. The power inequality is the ratio of the average degree in the minority group to the average degree in the majority group. The moment glass ceiling is the ratio of the second moment of degree distribution in the minority group to the second moment of degree distribution in the majority group. For Power Inequality and Moment Glass Ceiling, the minority nodes have the most advantage in Diversified Homophily BA, while the majority nodes have the most advantage in Homophily BA.

The color of each cell shows $\Delta I(t/T)$. The white color denotes $\Delta I(t/T) = 0$ (i.e., equality). The red color denotes $\Delta I(t/T) > 0$ (i.e., the majority group is at an advantage), while the blue color denotes $\Delta I(t/T) < 0$ (i.e., the minority group is at an advantage). We omit the performances under mid minority seeding because we found they are very close to performance under high seeding portion. The results for mid minority seeding are in the Additional file 1.

Under low minority seeding portion (less than 3 seeds, which is 0.3% in minority group), the majority group initially has an advantage ($\Delta I(t/T) > 0$) for all models, and subsequently converges to equality ($\Delta I(t/T) = 0$). Comparing the models, we observe that Diversified Homophily BA and its variation take shorter time to reach equality, while Homophily BA and Random Homophily take longer time. When the minority seeding

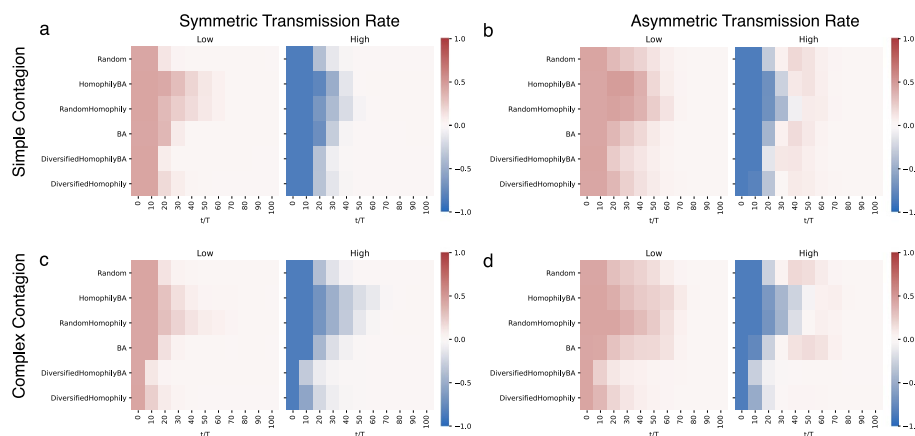


Fig. 2 Information spreading equality of different network models. We plot the spreading equality with low and high minority seeding portions. Each plot is a heatmap, where the x-axis represents the relative time t/T , the y-axis represents the different models, and the color represents $\Delta I(t/T)$. Recall that $\Delta I(t/T) = 0$ represents equality. When the minority seeding portion is low, $\Delta I(t/T)$ is initially positive (i.e., the majority group has a greater advantage); it then decreases to 0 for all models. When the minority seeding portion is high, $\Delta I(t/T)$ is initially negative (i.e., the minority group has a greater advantage). Under symmetric transmission, $\Delta I(t/T)$ increases to 0. However under asymmetric transmission, $\Delta I(t/T)$ increases to 0, then becomes positive, and eventually becomes 0. We observe that regardless of the contagion type and the seeding condition, Homophily BA and Random Homophily take longer to reach $\Delta I(t/T) = 0$, indicating less equality. On the contrary, Diversified Homophily BA and Diversified Homophily reach $\Delta I(t/T) = 0$ faster, sometimes even faster than Random Network and BA, indicating more equality. The differences between the models are more pronounced under complex contagion

portion is high (more than 7 seeds, which is 0.7% in minority group), the minority group initially has an advantage ($\Delta I(t/T) < 0$) and then converges to equality ($\Delta I(t/T) = 0$). Under asymmetric transmission, majority group has an advantage ($\Delta I(t/T) > 0$) in the middle of the process for some models. We again observe that Diversified Homophily BA and its variation take shorter time to reach equality, while Homophily BA and Random Homophily take longer time.

In summary, Homophily BA and Random Homophily achieve the lowest information access equality, followed by Random Network, BA and Diversified Homophily BA and their variation. The ranking between Random Network, BA, and Diversified Homophily BA depends on the process. Although Random Network and BA are the most equal in degree equality measures and have the largest heterophilicity values, they are not always the most equal in information access equality (e.g., under complex contagion and asymmetric transmission rate).

Experiment 2: effects of different network parameters on information access equality

We investigate the information access equality of network models with different parameters. Specifically, we investigate Homophily BA under different homophily h , preferential attachment strength α , and minority portion m . We also investigate Diversified Homophily BA under different diversification probability p_d .

What is the influence of homophily on information access equality? Figure 3 shows the impact of homophily on information access equality. We vary homophily h in the range between 0.5 (random mixing) and 1 (perfectly homophilic) under Homophily BA (with $m = 0.2, l = 2, \alpha = 1$). At low minority seeding portion, we observe that for

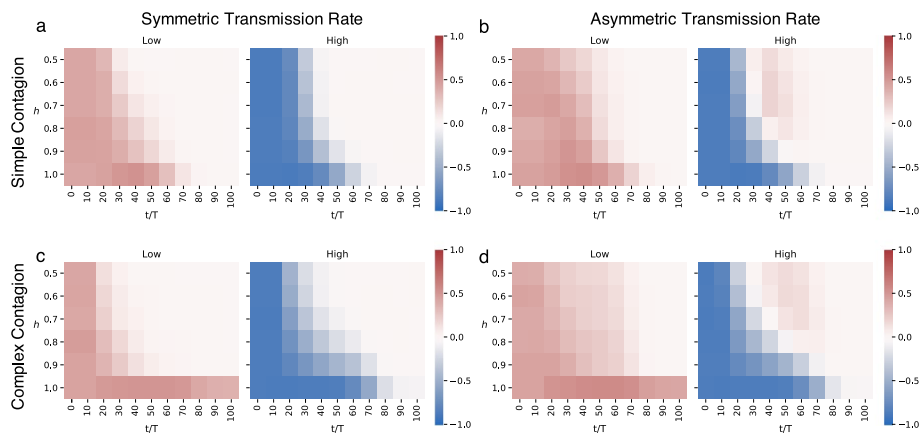


Fig. 3 Information spreading equality for homophily BA with different levels of homophily h . We plot the spreading equality for Homophily BA networks with low and high minority seeding portions. Each plot is a heatmap, where the x-axis represents the relative time t/T , the y-axis represents the graph's homophily h , and the color represents $\Delta I(t/T)$. Recall that $\Delta I(t/T) = 0$ represents equality. At low minority seeding, the lower h values, the more information access equality. This holds across all processes. However at high minority seeding portion, the behavior depends on the symmetry of the transmission. Under symmetric transmission rates (shown in **a** and **c**) with high minority seeding portion, lower h values achieve more information access equality. Under asymmetric transmission rates (shown in **b** and **d**) with high minority seeding portion, lower h values increase to zero, then become positive, before finally decreasing to zero. This makes it hard to judge which h value achieves more equality when one has an asymmetric transmission and high minority seeding

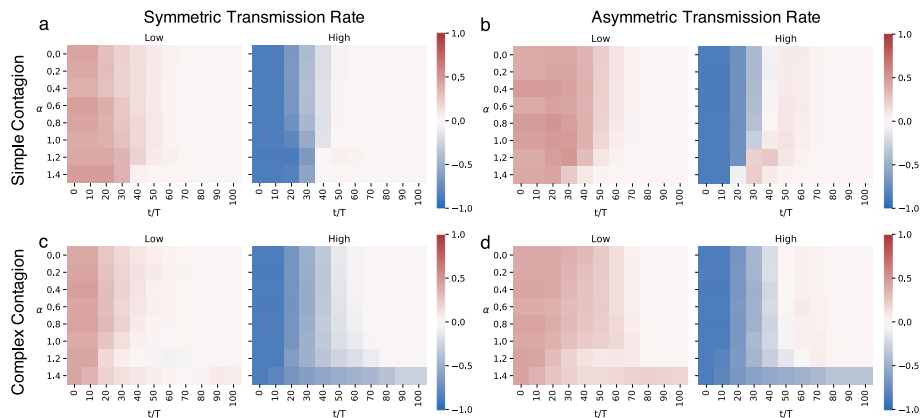


Fig. 4 Information spreading equality for homophily BA with different preferential attachment strengths α . When $\alpha < 1$, we observe few differences in information equality, implying that for sublinear preferential attachment, other factors such as homophily are more important for information access equality. However, for $\alpha \geq 1$ (which is often the case in real-world networks), the type of contagion is important. For example when $\alpha = 1.4$, simple contagion achieves information access equality quickly (shown in **a** and **b**). This is not true under complex contagion (shown in **c** and **d**). This highlights the importance of the type of contagion on the network

all processes smaller h values have more information access equality. At high minority seeding portion, under symmetric transmission rate (Fig. 3a, c), smaller h values also achieve more information access equality. However, under asymmetric transmission rate and high minority seeding portion, (Fig. 3b, d), we observe that the majority nodes gain an advantage around $t/T = 40$ and keep that advantage until $t/T = 70$ before equality is established. *What is the influence of preferential attachment strength on information access equality?* We experiment with α between 0 and 1.4, from no

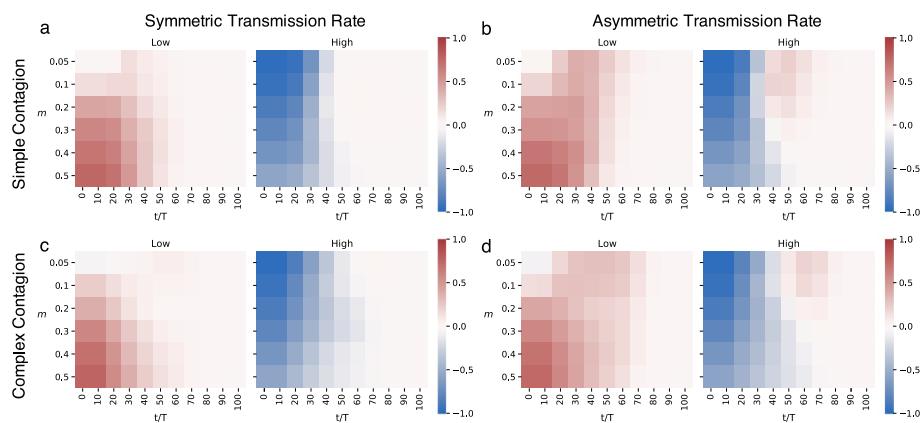


Fig. 5 Information spreading equality for homophily BA with different minority portions m . The influence of m on information access equality is strongly dependent on the minority seeding portion and is independent of contagion type and transmission rate. At low minority seeding portion, more equality is achieved with the lowest m (0.05 in our experiments)

preferential attachment ($\alpha = 0$) to preferential attachment ($\alpha > 1$). Figure 4 shows the impact of α on information access equality. When $\alpha < 1$, we find limited differences in information access equality, implying that with sublinear preferential attachment, other factors such as homophily are more important for information access equality. However, for $\alpha \geq 1$, under simple contagion (Fig. 4a, b) we observe a slight decrease and then an increase in equality. $\alpha = 1.4$ achieves the most information access equality. However, under complex contagion (Fig. 4c, d), we find that $\alpha = 1.4$ achieves the least information access equality. This could be related to the different functions hubs play under simple and complex contagion. Under simple contagion, hubs are pathways, while under complex contagion, hubs are bottlenecks. This observation suggests that the occurrence of hubs on information access equality seems to be significant only under super-linear preferential attachment.

What is the influence of minority portion on information access equality? For the minority portion m , we experiment in the range between 0.05 (almost no minority nodes) and 0.5 (no population difference between majority and minority nodes). We find that the influence of m on information access equality is strongly dependent on minority seeding portion; this is independent of contagion type and transmission rate (see Fig. 5). At low minority seeding portion, more equality is achieved with the smallest m (0.05 in our experiments). At high seeding proportion, the larger m values achieve more equality (0.5 under our experiment). This can be explained by the fact that minority seeding portion (as an absolute value), has a different relative effect among different minority populations. For example, minority seeding portion of 0.3 is considered smaller when $m = 0.5$ compared to $m = 0.1$.

What is the influence of diversity on information access equality? We investigate the behavior of Diversified Homophily BA with different p_d values: 0.01, 0.05, 0.1, 0.2, 0.4, 0.6 and 0.8. We set $m = 0.2$, $h = 0.8$, $l_d = 1$, and $\alpha = 1$. We observe that, in general, larger p_d values achieve more information access equality, especially for low minority seeding portion (Fig. 6). This matches our observation that larger p_d values have more degree equality and less homophily.

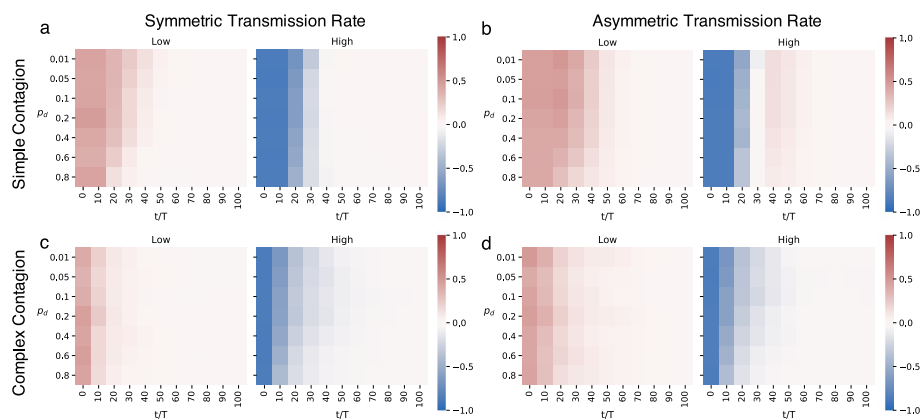


Fig. 6 Information spreading equality for diversified homophily BA with different diversification probabilities p_d . In general, we observe that the higher the p_d , the higher the information access equality, especially for low minority seeding portion

Experiment 3: information access equality in real-world networks

We experiment with three real-world networks: (1) The Github follower network, (2) The DBLP collaboration network, and (3) The APS citation network (Lee et al. 2019). The Github and DBLP networks have gender as a grouping attribute. APS has research field as the grouping attribute. We filter the networks by including only labeled nodes and selecting the largest connected component. The basic statistics of the three networks are in the table presented as part of Fig. 7b. For these information access simulations, we keep the parameters the same as in Experiments 1 and 2, except for the seeding portion and the number of simulations. We run experiments on 0.2%, 1% and 10% seeding portions. Since the results across these seeding portions are similar, we only report the results on 0.2% seeding portion here. The figures for the other experiments are in the Additional file 1. For APS, 0.2% of the nodes is only 2 nodes, which makes it difficult to enforce a different portion of minority seeding, so we set $s = 5$. The results reported below are averages over 100 simulations.

Figure 7a shows the degree distributions of the real networks. We observe that all networks have heavy-tailed degree distributions. The basic statistics for these networks are in Fig. 7b. We note that all networks have dyadicity greater than 1 and heterophilicity less than 1, indicating homophily in the network. Comparing across the networks, we observe that APS has the largest dyadicity and the smallest heterophilicity, followed by Github and DBLP. The social capital (degree) equality measures are in Fig. 7c. In terms of degree equality, Github is the most equal, followed by DBLP and then APS.

Figure 8 shows information access equality on real networks. For Github, we find that initially $\Delta I(t/T) < 0$, indicating that minority nodes have a greater advantage. This is because Github has the smallest minority portion, and this is consistent with the observation for influence on equality with different minority portion m values (see Fig. 5). Across all process settings, we find that APS takes the longest to reach equality. This is consistent with our previous observation that APS is less equal in degree and has more homophily. DBLP is most equal in information access, which is consistent with less homophily. We find that the information access equality landscape depends on different process settings. For example, under asymmetric transmission rate, we notice that

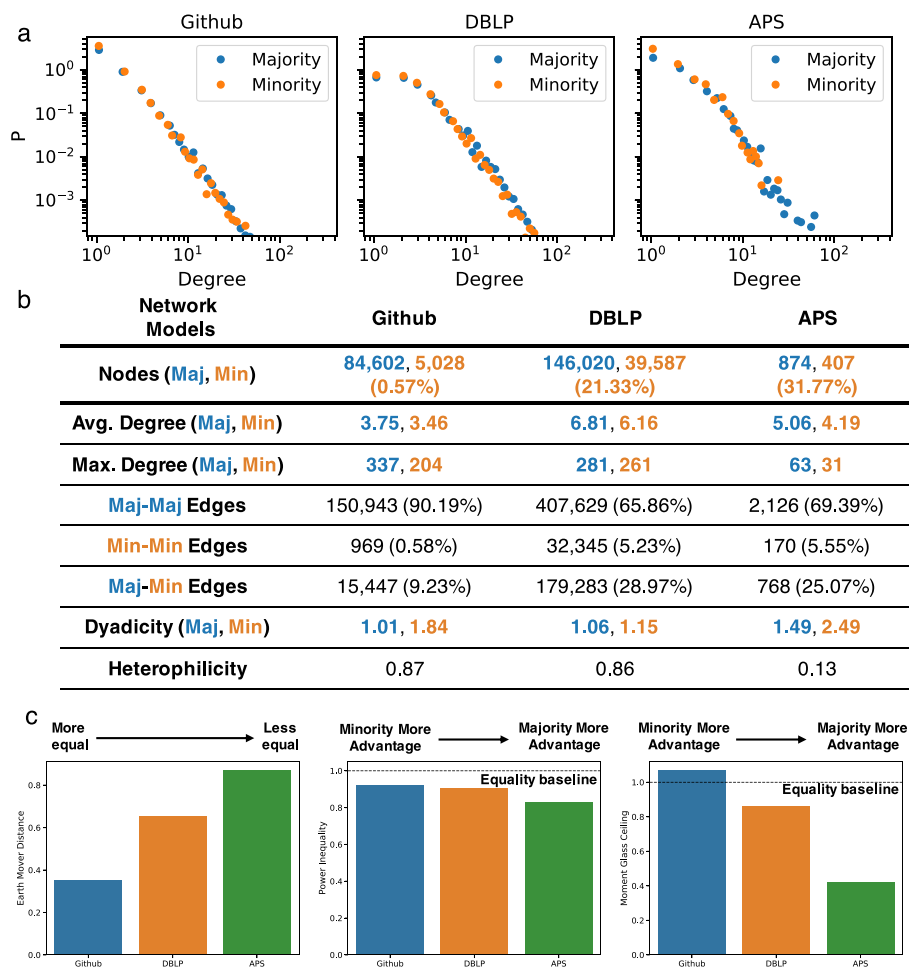


Fig. 7 Statistics of real networks. **a** Degree Distributions. We see that all networks have heavy-tail degree distributions, but we note that the minority group in APS has a much shorter tail. **b** Basic Statistics. The table lists basic statistics, including node composition, average and maximum degree, edge composition, dyadicity and heterophilicity scores for the three networks. Github has the lowest proportion of minority nodes, while APS has the highest proportion of minority nodes. DBLP has the highest average and maximum degree. All the networks have dyadicity greater than 1 and heterophilicity less than 1, indicating homophilic behavior. Among them, APS has the highest dyadicity value and the lowest heterophilicity value. **c** Social Capital Equality Measures. The Earth Mover Distance of the three networks are all positive, indicating that the degree distribution is different between the groups. APS has the highest Earth Mover Distance. All networks have power inequality less than 1, indicating that minority groups have a lower average degree. For the moment glass ceiling, GitHub is slightly higher than 1, but the other two networks both have a moment glass ceiling lower than 1, indicating that there are fewer minority nodes as hubs

$\Delta I(t/T)$ becomes positive for Github and DBLP under high seeding portion, which is not the case under symmetric transmission rate. We also find that achieving equality is much harder under complex contagion and asymmetric transmission rate. These variations are consistent with our earlier observation that equality in information access is related to both the network structure and the properties of the spreading process (see Fig. 8).

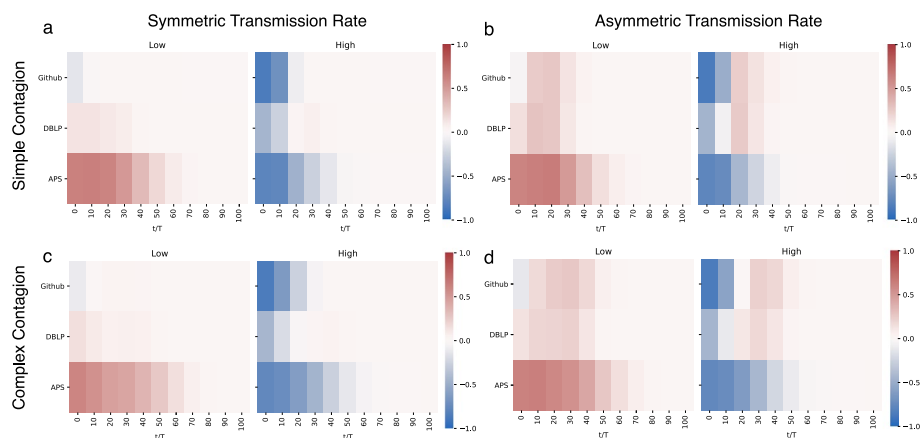


Fig. 8 Information spreading equality of real networks. Recall that we plot the spreading equality with low and high minority seeding portions. Each plot is a heatmap, where the x-axis represents the relative time t/T , the y-axis represents the different real-world networks, and the color represents $\Delta I(t/T)$ while $\Delta I(t/T) = 0$ represents equality. For Github, we find that initially $\Delta I(t/T) < 0$ for high minority seeding portion and $\Delta I(t/T) \approx 0$ for low minority seeding portion, indicating that majority nodes don't have many advantage. This is because Github has the smallest minority portion, and this is consistent with the observation for influence on equality across different m values (see Fig. 5). Across all process settings, we find that APS takes the longest to reach equality, which is consistent with our previous observation that APS is less equal in degree and has a higher homophily level. DBLP is the most equal in information access, which is consistent with lower homophily. We find that the information access equality landscape depends on different process settings. For example, under asymmetric transmission rate, we notice that $\Delta I(t/T)$ becomes positive for Github and DBLP under high seeding portion, which is not the case under symmetric transmission rate. We also find that achieving equality is much harder under complex contagion and asymmetric transmission rate

Experiment 4: regression analysis to detect significant features in information access equality

To systematically analyze what features contributes to information access equality, we construct datasets with multiple network features and process features, and build linear regression models to predict $\Delta I(t/T)$ at $t/T = 40$. We perform feature scaling on the numerical features and one-hot encoding on the categorical features. We observe that regardless of the experiment (namely, Experiments 1, 2, or 3), the process features are all statistically significant at p -value of 0.05. We observe that a symmetric transmission rate will increase equality compared to an asymmetric one. A simple contagion will decrease equality compared to a complex contagion. Compared to a mid seeding portion, a low minority seeding portion will decrease equality and a high minority seeding portion will increase equality. However, all network features are not statistically significant across all of the experiments. We observe that degree inequality (as measured by power inequality or moment glass ceiling) is significant in 80% of the experiments. The higher the power inequality, the higher the information access inequality. This means that when minority nodes have more advantage, the information access inequality is lower. We also observe that network distance (as measured by shortest path or diameter) is significant in 60% of the experiments. Additionally, the higher the (Min, Min) shortest path, the more information access inequality, and the lower the (Maj, Min) shortest path, the more information access equality. For brevity, we do not include the full regression table here; however, the reader will find the table in the Additional file 1.

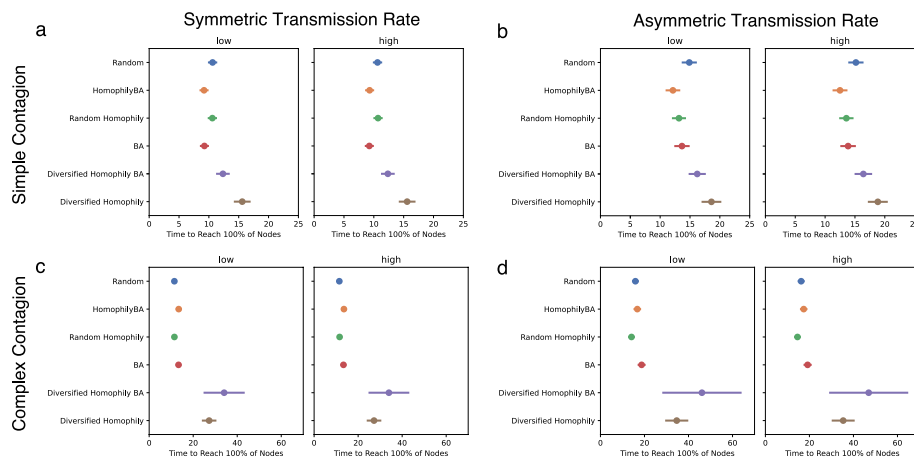


Fig. 9 Information spreading efficiency of different generative models of complex networks. We inspect information access efficiency by looking at the time needed to reach 100% of the nodes. The x-axis is the time (in the number of iterations) needed to reach 100% of the nodes. The y-axis is the different generative models. We show the average and standard deviation values of the time needed for 100% outreach across 100K runs for each model and each spreading process. On average, more time is needed with asymmetric transition. Diversified Homophily and Diversified Homophily BA are, on average, slower in spreading information than other models (especially with a complex contagion). This is due to their inter-group edges. With complex contagion, the standard deviation on time for 100% outreach is larger for Diversified Homophily BA. We conjecture that this is due to the fact that hubs become bottlenecks in complex contagion and Diversified Homophily BA enforces diversification. The hyperparameters used here are the same as in Experiment 1: $p_d = 0.6$, $h = 0.8$, $\alpha = 1$, and $m = 20\%$. For a heatmap view of efficiency in information spreading of different generative models, see the Additional file 1.

Experiment 5: relationship between information equality and spreading efficiency

Research on fairness in AI has shown a phenomenon called the “price of fairness” (Menon and Williamson 2018; Corbett-Davies et al. 2017), where researchers have found that fairness comes at a cost to other performance measures of interest. For information spreading, the relationship between equality and efficiency is a rather important one, especially when the information is time-sensitive. Past work has found the “price of fairness” under information maximization (Tsang et al. 2019).

Figure 9 shows the efficiency of information spread across different network models. The x-axis is the time (in the number of iterations) needed to reach 100% of the nodes. The y-axis is the different generative network models. We show the average and standard deviation values of the time needed to reach 100% of the nodes across 100K runs for each model and each spreading process. On average, more time is needed with asymmetric transition. Diversified Homophily and Diversified Homophily BA are, on average, slower in spreading information than other models because they have more inter-group edges. Recall that the same inter-group edges slow down the spreading process help increase information access equality. The spreading inefficiency is more pronounced with complex contagion. With complex contagion, the standard deviation on time to reach 100% of the nodes is larger for Diversified Homophily BA. This is most likely due to the fact that hubs become bottlenecks in complex contagion and Diversified Homophily BA enforces diversification.⁴ For a heatmap view of efficiency in information spread across different generative models, see the Additional file 1.

⁴ The hyperparameters used here are the same as in Experiment 1: $p_d = 0.6$, $h = 0.8$, $\alpha = 1$, and $m = 20\%$.

By inspecting different network parameters, we observe that except for the minority population m , all other parameters can affect the information spreading efficiency. For brevity, we report a summary of these results here. The Additional file 1 include the relevant figures. For homophily h , $h = 1$ always has the lowest efficiency, possibly due to the extreme case where only one edge connects the two groups. We observe that for asymmetric transmission rate, a lower h has a slightly slower spreading speed due to more inter-group edges, again showing the equality and efficiency trade-off of inter-group edges.

We observe a clear trend in spreading efficiency and the preferential attachment strength α . Under simple contagion, the larger the α , the faster the spreading. In contrast, under complex contagion, the larger the α , the slower the spreading speed. This is to be expected because the emergence of hubs promotes simple contagion but hinders complex contagion. Recall that we observed that larger α values have higher equality in simple contagion and lower equality in complex contagion.

We observe a slight decrease in spreading efficiency as diversification probability p_d varies under simple contagion. Under complex contagion, larger p_d values significantly decrease the spreading speed, again showing the trade-off between efficiency and equality with increasing inter-group edges.

In short, we find that there is a trade-off between information access equality and information spreading efficiency. This is mainly due to the inter-group edges. However, similar to the observation that equality is related to spreading process setting, this trade-off is also dependent on the spreading process setting.

Conclusion and discussion

We focused on information access equality in complex networks when the population is divided into two mutually exclusive groups: majority vs. minority. We measured the information access equality of various processes on different network models with different parameters and on three real networks. We found that information access equality depends not only on the network structure, but also on different contagion types (i.e., simple vs. complex), different transmission rates (symmetric vs. asymmetric), and different minority proportions. For instance, the occurrence of hubs on information access equality seems to be significant only under super-linear preferential attachment. Based on regression analyses, we observed that process features are all statistically significant (p -value ≤ 0.05) for information access equality. A symmetric transmission rate will increase equality compared to an asymmetric one. A simple contagion will decrease equality compared to a complex contagion. Compared to a mid seeding portion, a low minority seeding portion will decrease equality and a high minority seeding portion will increase equality. The network features were not always statistically significant. However, two network features stood out: degree inequality (as measured by power inequality or moment glass ceiling) and network distance (as measured by shortest path or diameter). Although it is very difficult to draw a single conclusion about what type of networks can promote information access equality, we find that, in general, more inter-group edges can help achieve equality. However, we note one drawback to more inter-group edges: they may reduce the efficiency of information spreading. Designing a network with information access equality requires more knowledge about the spreading process itself. Our findings can be used to guide recommendations for mechanistic design of social

networks that foster more equal information access. For example, social networking platforms, such as Facebook and LinkedIn, can recommend new connections that lead to more equal information access for various spreading processes on their platforms. Ferrara et al. Ferrara et al. (2022) show how link recommendation algorithms can reduce the visibility of the minority group when both groups are heterophilic. Deployment of our findings in a real-world system is part of our future work.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1007/s41109-022-00494-8>.

Additional file 1. Supplementary information for additional experiment results including network measure of models with different parameters, information access efficiency, linear regression analysis on information access equality and information access equality on real networks with different seeding population.

Acknowledgements

TER acknowledges her Volkswagen Foundation Grant on “Reclaiming Individual Autonomy and Democratic Discourse Online” for shaping her initial thoughts on this work.

Author contributions

XW, OV and TER designed the research and prepared the paper. XW conducted the experiments and performed numerical analysis of the data. All authors approved the manuscript.

Funding

The authors were not supported by external funding for this work.

Availability of data and materials

The datasets generated and/or analysed during the current study are available in the NetInfoAccEqua repository: <https://github.com/xindi-dumbledore/NetInfoAccEqua>.

Declarations

Competing interests

The authors declare that they have no competing interests.

Received: 8 December 2021 Accepted: 20 July 2022

Published online: 02 August 2022

References

- Adamic LA, Glance N (2005) The political blogosphere and the 2004 US election: Divided they blog. In: *LinkKDD*, pp 36–43
- Aiello LM, Deplano M, Schifanella R, Ruffo G (2012) People are strange when you're a stranger: impact and influence of bots on social networks. In: *ICWSM*
- Albert R, Barabási A-L (2002) Statistical mechanics of complex networks. *Rev Modern Phys* 74(1):47
- Ali J, Babaei M, Chakraborty A, Mirzasoleiman B, Gummadi KP, Singla A (2019) On the fairness of time-critical influence maximization in social networks. *arXiv preprint arXiv:1905.06618*
- Anderson RM, May RM (1992) *Infectious diseases of humans: dynamics and control*. Oxford University Press, Oxford, England
- Aral S, Dhillon PS (2018) Social influence maximization under empirical influence models. *Nat Hum Behav* 2(6):375–382
- Aral S, Walker D (2011) Creating social contagion through viral product design: a randomized trial of peer influence in networks. *Manag Sci* 57(9):1623–1639
- Avin C, Keller B, Lotker Z, Mathieu C, Peleg D, Pignolet YA (2015) Homophily and the glass ceiling effect in social networks. In: *ITCS*, pp 41–50
- Barberá P, Jost JT, Nagler J, Tucker JA, Bonneau R (2015) Tweeting from left to right: is online political communication more than an echo chamber? *Psychol Sci* 26(10):1531–1542
- Barrat A, Barthelemy M, Vespignani A (2008) *Dynamical processes on complex networks*. Cambridge University Press, Cambridge, England
- Becker R, D'Angelo G, Ghobadi S, Gilbert H (2021) Fairness in influence maximization through randomization. In: *AAAI*, pp 14684–14692
- Berlingerio M, Koutra D, Eliassi-Rad T, Faloutsos C (2013) Network similarity via multiple social theories. In: *ASONAM*, pp 1439–1440
- Boyd D, Levy K, Marwick A (2014) The networked nature of algorithmic discrimination. *Collected Essays, Data and Discrimination*
- Burt RS (2000) The network structure of social capital. *Res Org Behav* 22:345–423
- Calvo-Armengol A, Jackson MO (2004) The effects of social networks on employment and inequality. *Am Econ Rev* 94(3):426–454

- Centola D (2010) The spread of behavior in an online social network experiment. *Science* 329(5996):1194–1197
- Centola D, Macy M (2007) Complex contagions and the weakness of long ties. *Am J Soc* 113(3):702–734
- Corbett-Davies S, Pierson E, Feller A, Goel S, Huq A (2017) Algorithmic decision making and the cost of fairness. In: KDD, pp 797–806
- Daley DJ, Kendall DG (1964) Epidemics and rumours. *Nature* 204(4963):1118–1118
- DiMaggio P, Cohen J (2021) Information inequality and network externalities: a comparative study of the diffusion of television and the internet. In: *The Economic Sociology of Capitalism*, pp 227–267. Princeton University Press, Princeton, NJ
- Erdős P, Rényi A (1960) On the evolution of random graphs. *Publ Math Inst Hung Acad Sci* 5(1):17–60
- Ferrara A, Espín-Noboa L, Karimi F, Wagner C (2022) Link recommendations: their impact on network structure and minorities. In: *WebSci*
- Fish B, Bashardoust A, Boyd D, Friedler S, Scheidegger C, Venkatasubramanian S (2019) Gaps in information access in social networks? In: *WWW*, pp 480–490
- Freese J, Lutfey K (2011) Fundamental causality: challenges of an animating concept for medical sociology. In: Pescosolido BA, Martin JK, McLeod JD, Rogers A (eds) *Handbook of the sociology of health, illness, and healing: a blueprint for the 21st century*. Springer, New York, NY, pp 67–81
- Garrett RK (2009) Echo chambers online?: Politically motivated selective exposure among internet news users. *J Comput Mediat Commun* 14(2):265–285
- Gentzkow M, Shapiro JM (2011) Ideological segregation online and offline. *Q J Econ* 126(4):1799–1839
- Granovetter MS (1973) The strength of weak ties. *Am J Sociol* 78(6):1360–1380
- Granovetter M (1978) Threshold models of collective behavior. *Am J Sociol* 83(6):1420–1443
- Holland PW, Laskey KB, Leinhardt S (1983) Stochastic blockmodels: first steps. *Soc Netw* 5(2):109–137
- Hymowitz C, Schellhardt TD (1986) The glass ceiling: why women can't seem to break the invisible barrier that blocks them from the top jobs. *Wall Str J* 24(1):1573–1592
- Iyengar S, Hahn KS (2009) Red media, blue media: evidence of ideological selectivity in media use. *J Commun* 59(1):19–39
- Jalali ZS, Wang W, Kim M, Raghavan H, Soundarajan S (2020) On the information unfairness of social networks. In: *SDM*, pp 613–521
- Karimi F, Génois M, Wagner C, Singer P, Strohmaier M (2018) Homophily influences ranking of minorities in social networks. *Sci Rep* 8(1):1–12
- Kempe D, Kleinberg J, Tardos É (2003) Maximizing the spread of influence through a social network. In: *KDD*, pp 137–146
- Lee E, Karimi F, Wagner C, Jo H-H, Strohmaier M, Galesic M (2019) Homophily and minority-group size explain perception biases in social networks. *Nat Hum Behav* 3(10):1078–1087
- Menon AK, Williamson RC (2018) The cost of fairness in binary classification. In: *FACCT*, pp 107–118
- Messias J, Schmidt L, Oliveira R, Benevenuto F (2013) You followed my bot! Transforming robots into influential users in twitter. *First Monday*
- Messias J, Vikatos P, Benevenuto F (2017) White, man, and highly followed: gender and race inequalities in twitter. In: *WI*, pp 266–274
- Mønsted B, Sapiężyński P, Ferrara E, Lehmann S (2017) Evidence of complex contagion of information in social media: an experiment using twitter bots. *PLoS One* 12(9):0184148
- Oliveira M, Karimi F, Zens M, Schaible J, Génois M, Strohmaier M (2021) Mixing dynamics and group imbalance lead to degree inequality in face-to-face interaction. *arXiv preprint arXiv:2106.11688*
- Park J, Barabási A-L (2007) Distribution of node characteristics in complex networks. *Proc Nat Acad Sci* 104(46):17916–17920
- Prior M (2007) *Post-broadcast democracy: how media choice increases inequality in political involvement and polarizes elections*. Cambridge University Press, Cambridge, England
- Rahmattalabi A, Jabbari S, Lakkaraju H, Vayanos P, Izenberg M, Brown R, Rice E, Tambe M (2021) Fair influence maximization: a welfare optimization approach. In: *AAAI*, pp 11630–11638
- Romero DM, Meeder B, Kleinberg J (2011) Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter. In: *WWW*, pp 695–704
- Shao C, Ciampaglia GL, Varol O, Yang K-C, Flammini A, Menczer F (2018) The spread of low-credibility content by social bots. *Nat Commun* 9(1):1–9
- Stoica AA, Chaintreau A (2019) Fairness in social influence maximization. In: *WWW*, pp 569–574
- Stoica AA, Han JX, Chaintreau A (2020) Seeding network influence in biased networks and the benefits of diversity. In: *WWW*, pp 2089–2098
- Stoica AA, Riederer C, Chaintreau A (2018) Algorithmic glass ceiling in social networks: the effects of social recommendations on network diversity. In: *WWW*, pp 923–932
- Tsang A, Wilder B, Rice E, Tambe M, Zick Y (2019) Group-fairness in influence maximization. In: *IJCAI*, pp. 5997–6005
- Varol O, Uluturk I (2020) Journalists on twitter: self-branding, audiences, and involvement of bots. *J Comput Soc Sci* 3(1):83–101
- Varol O, Ferrara E, Davis C, Menczer F, Flammini A (2017) Online human-bot interactions: detection, estimation, and characterization. In: *ICWSM*, pp 280–289
- Venkatasubramanian S, Scheidegger C, Friedler SA, Clauset A (2021) Fairness in networks: social capital, information access, and interventions. In: *KDD*, pp 4078–4079
- Vespignani A (2012) Modelling dynamical processes in complex socio-technical systems. *Nat Phys* 8(1):32–39
- Weng L, Menczer F, Ahn Y-Y (2013) Virality prediction and community structure in social networks. *Sci Rep* 3(1):1–6
- Woolley SC (2016) Automating power: social bot interference in global politics. In: *First Monday*
- Zhang Y, Han JX, Mahajan I, Bengani P, Chaintreau A (2021) Chasm in hegemony: explaining and reproducing disparities in homophilous networks. *arXiv preprint arXiv:2102.11925*

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.